Predicting What We Breathe

Jeanne Holm (PI, City of Los Angeles)
Dr. Mohammad Pourhomayoun (Co-I, California State University, Los Angeles)
Jeremy Taub (Co-I, OpenAQ)
Dawn Comer (Project Manager, City of Los Angeles)

ESTF, Atmospheric Gases - June 17, 2021

Background and Objectives

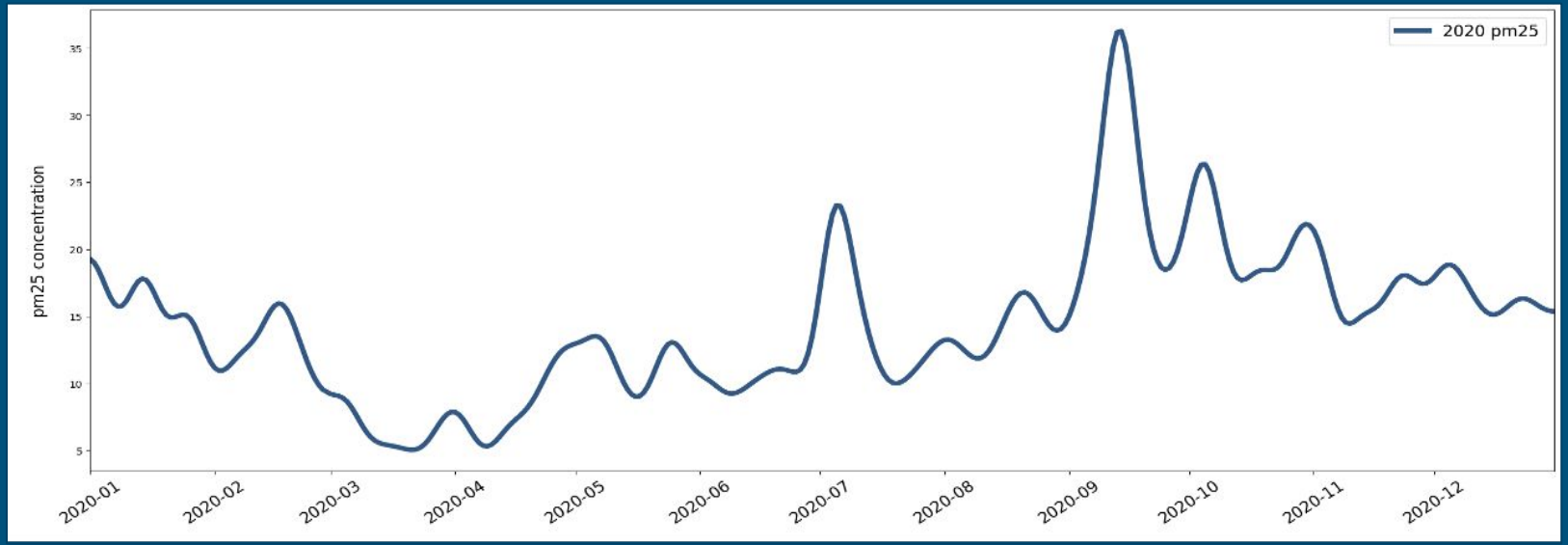
- **Project Objectives**
  - Increase accessibility and use of space data by using machine learning to help cities predict air quality in ways that will improve human health
  - Provide tools and algorithms to future Earth science missions (such as MAIA) to provide rapid ground truth, conduct data fusion across diverse datasets, and support rapid use of mission data
    - 1. Create a model for cities to enamine in-situ $PM_{2.5}$, NO_2, PM_{10}, and ozone
    - 2. Apply machine learning to big datasets from ground and space
    - 3. Improve decision making on health outcomes in cities

Background and Objectives

- **Approach**
  - Develop machine learning algorithms for predictive models for air quality based on measurements of $PM_{2.5}$ and other air pollutants
  - Develop a big data analytics algorithm for integrating ground and space data
  - Develop predictive models for health risks via deep learning, machine learning
  - Build an open source $PM_{2.5}$ stack for integrating ground and space data
  - Create a model for cities to understand predictions and effective interventions
  - Address issues of environmental justice in highly affected communities
  - Engage our community/citizen science volunteers

Air Pollution Prediction

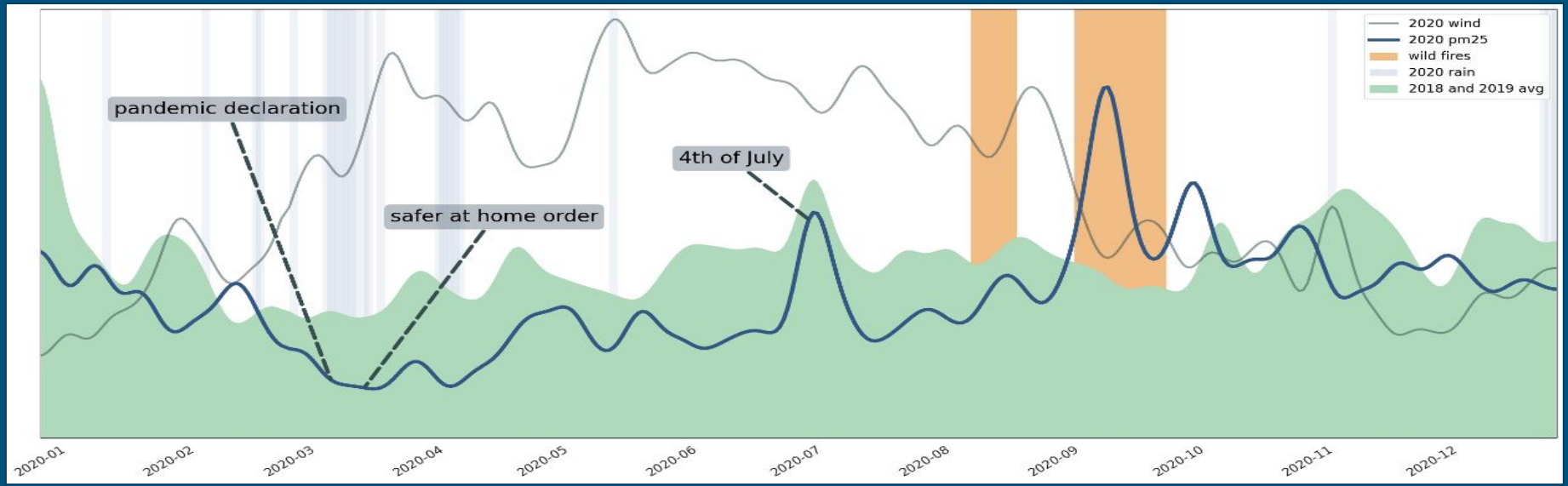
Collect data, process it, understand it, discover patterns, and predict air quality



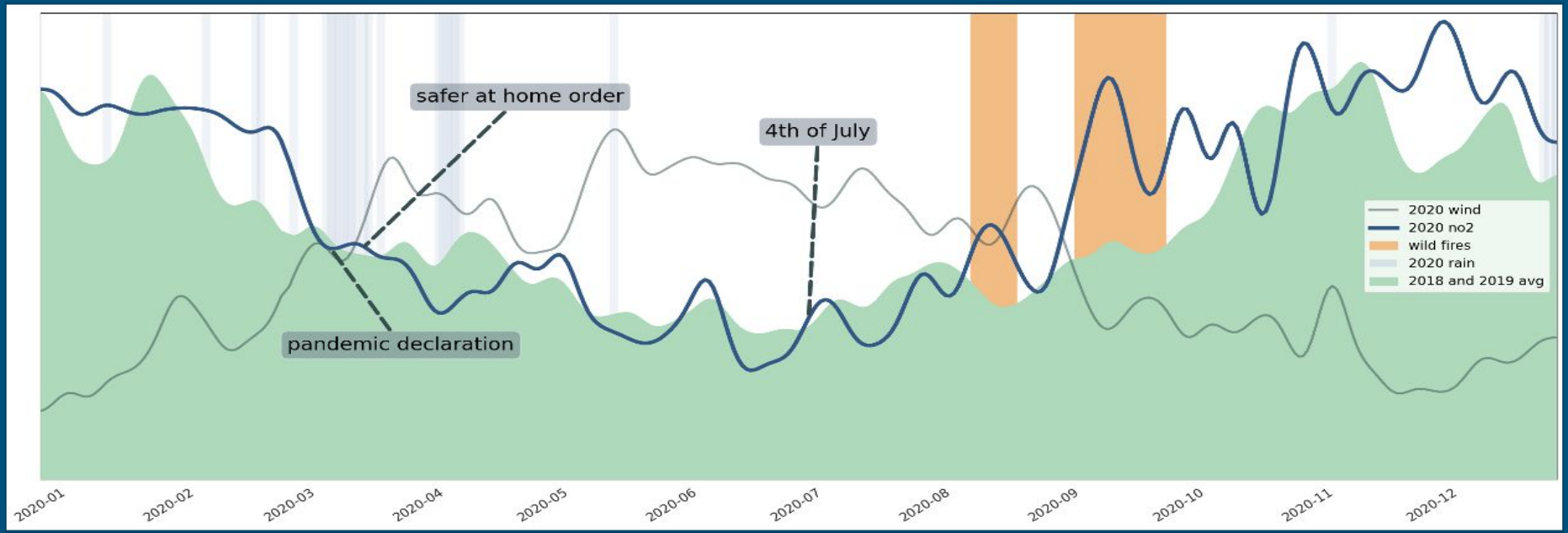
PM2.5 2020

Air Pollution Prediction - PM2.5

- Take into account all factors impacting air quality
- Collect, process, and use data from many sources
- Have a complex machine learning model to discover, extract, and learn patterns

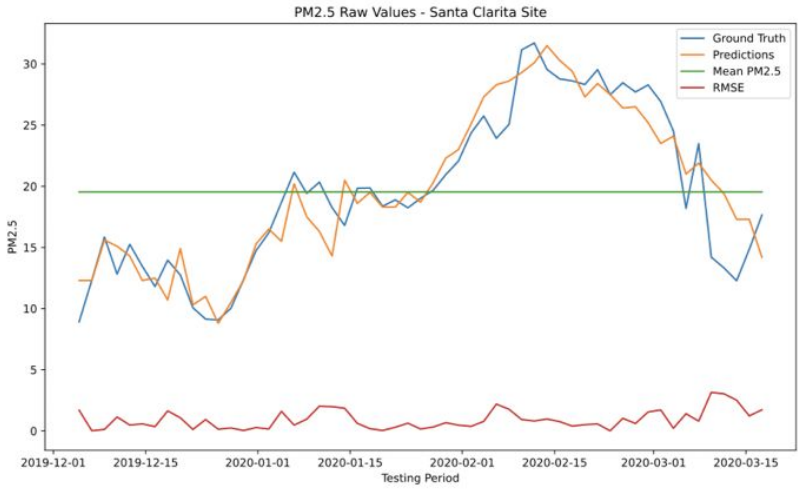


Air Pollution Prediction - NO2

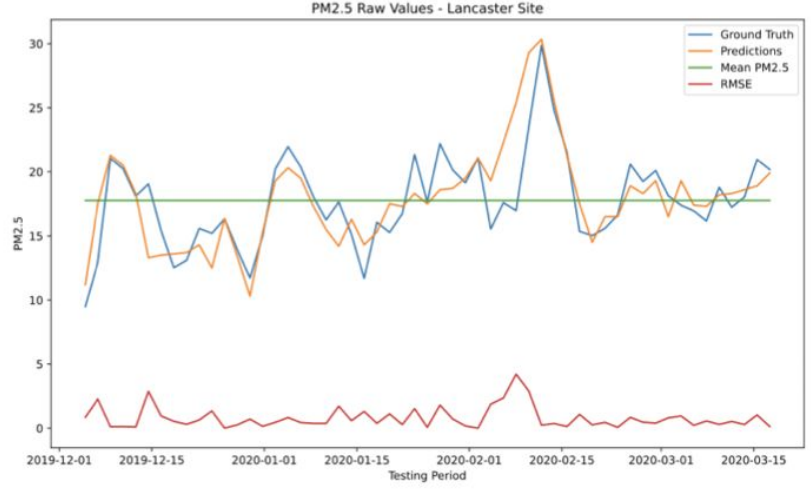


Sample Prediction Results for PM2.5: Satellite Images, Ground-Based Sensors, and Meteorological Data

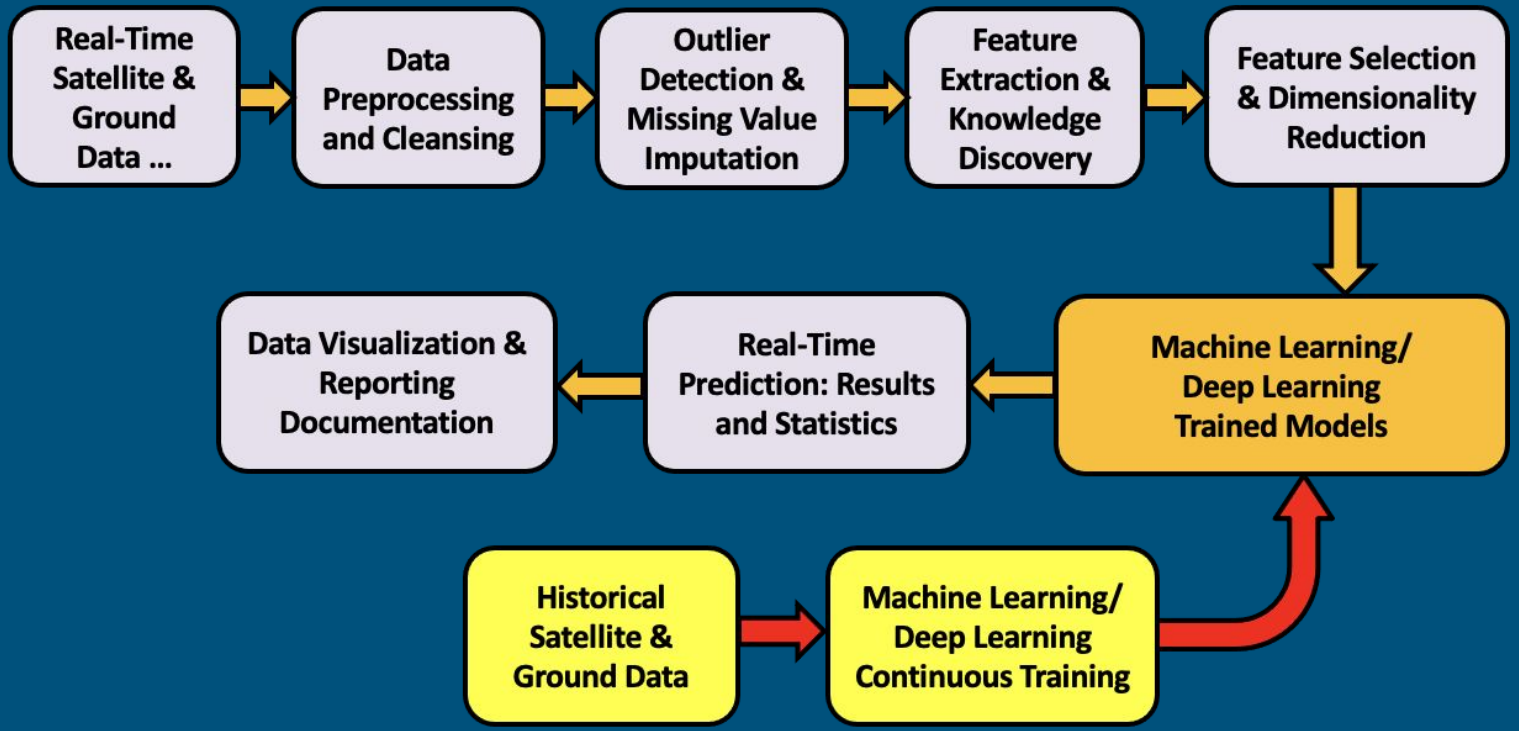
Santa Clarita Site PM 2.5 Observed Sensor Data vs Predicted



Lancaster Site PM 2.5 Observed Sensor Data vs Predicted

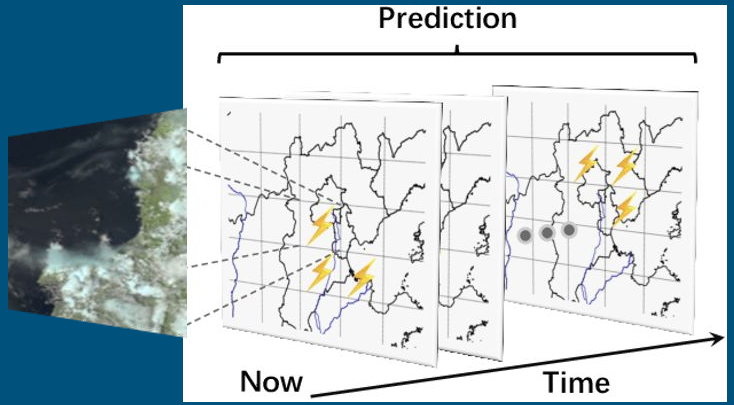


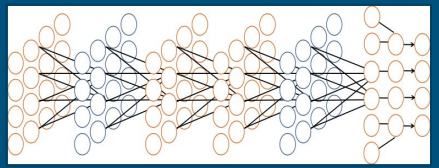
High-Level Structure for Predictive Models

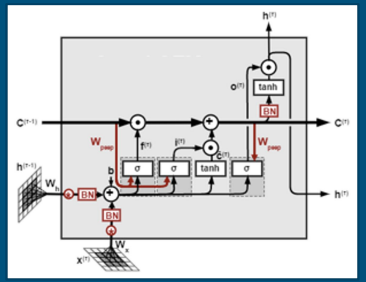


Machine Learning/Deep Learning Models

- **Deep Neural Networks**
- Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM): For the temporal correlation in the data
- Convolutional Neural Network (CNN): For the spatial correlation
- Combining deep learning models + algorithms = spatiotemporal correlation



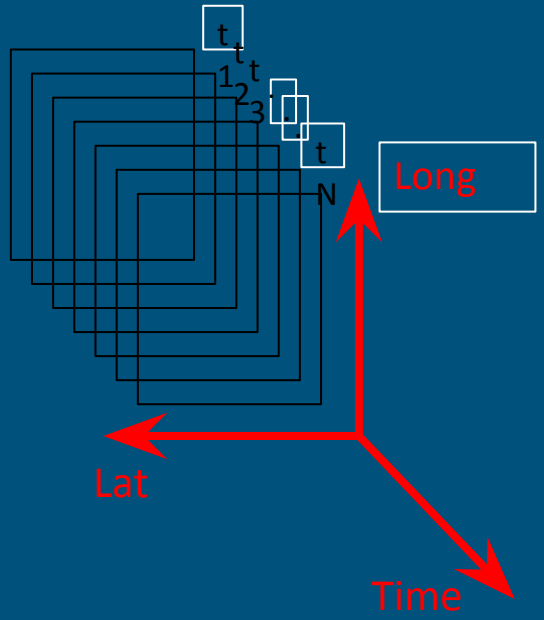




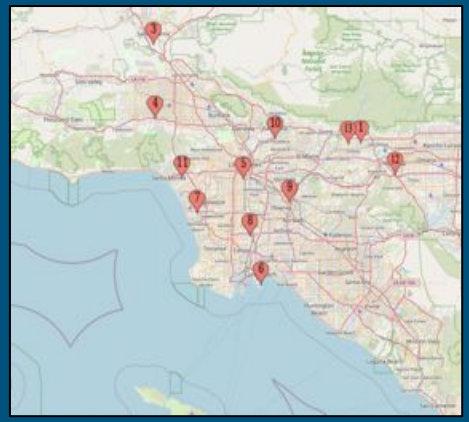
Temporal Resolution/Spatial Resolution

Considering Temporal and Spatial Patterns in the Data

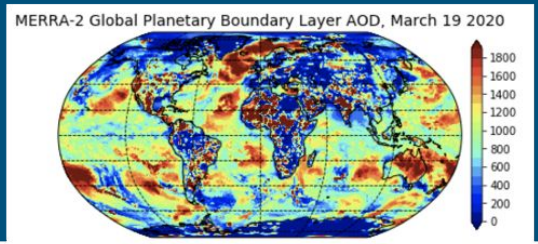
Temporal Correlation

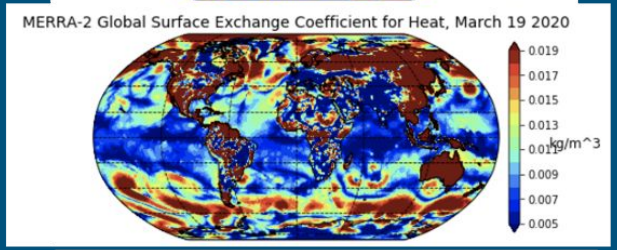


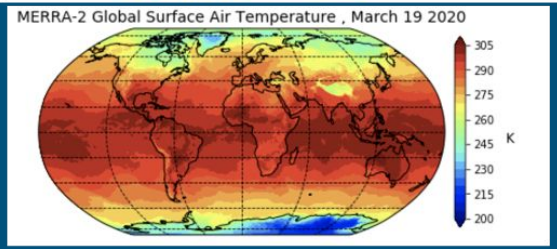
Spatial Correlation

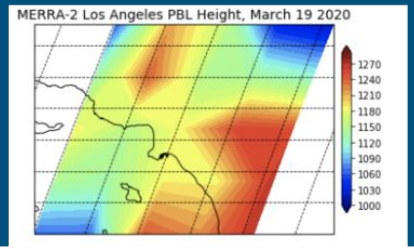


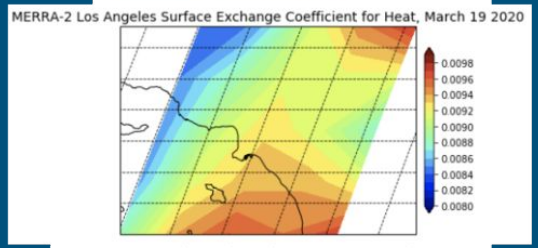
Wildfire/Smoke Data

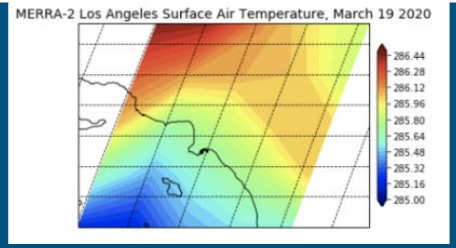


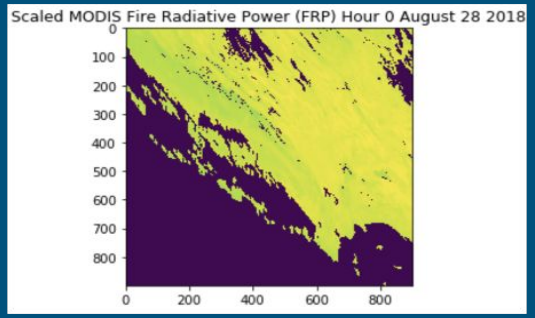


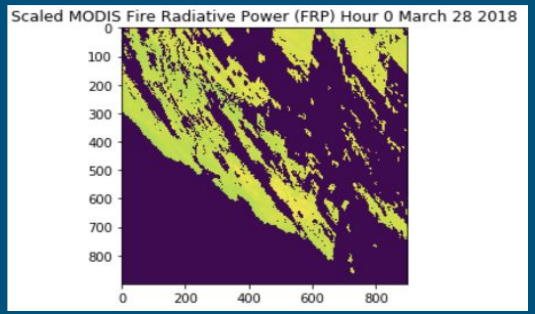




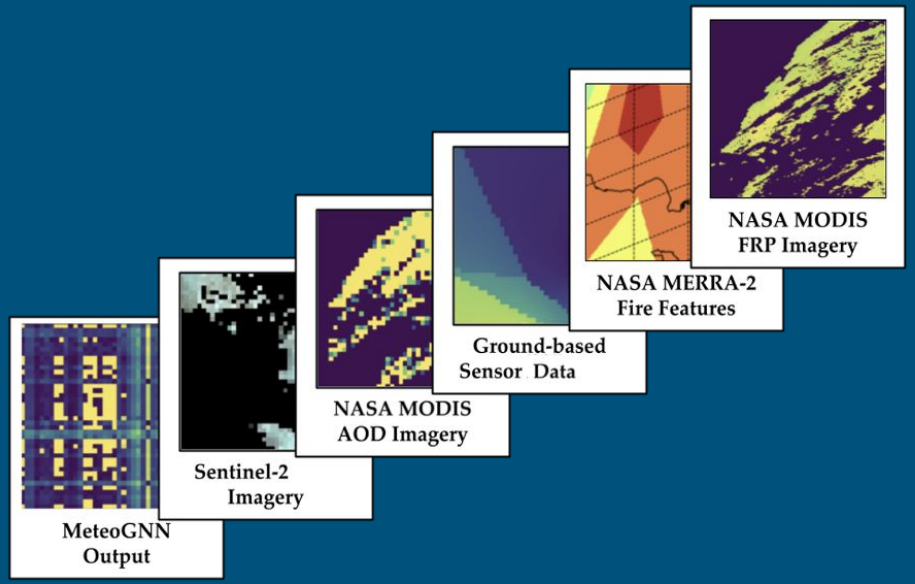








Predictive Model and Data



Data Processing and Data Fusion

- Preprocessing and cleansing
- Outliers/trustworthiness and missing values
- Feature extraction and knowledge discovery
- Feature selection and dimensionality Reduction
- Format matching and alignments

Predicting **PM2.5** Based on Satellite, Ground, Meteorological, and Wildfire/Smoke Data

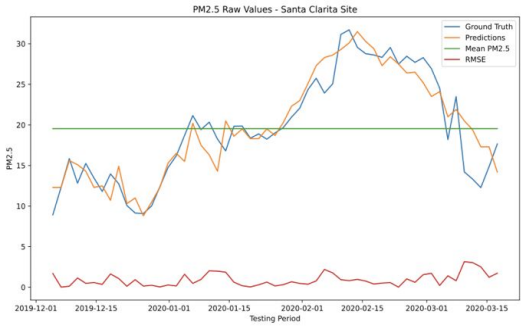
Input data

- Satellite observations NASA MODIS:
  - 1 km x 1 km/pixel
- Ground-based sensors (13 in L.A. County), hourly
- Wildfire/Smoke data from NASA MODIS, MERRA-2
- Meteorological data

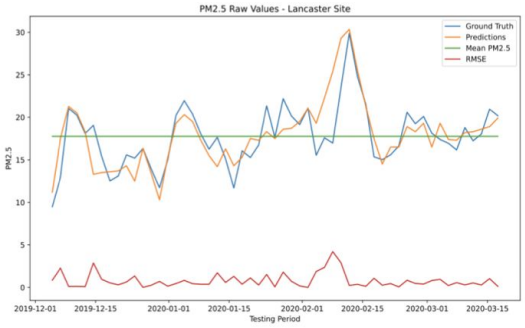
Accuracy	Frame #
93%	Frame 1: 2 days in future
90%	Frame 2: 4 days in future
88%	Frame 3: 6 days in future
83%	Frame 4: 8 days in future
80%	Frame 5: 10 days in future

Predicting **PM2.5** Based on Satellite, Ground, Meteorological, and Wildfire/Smoke Data

Santa Clarita Site PM 2.5 Observed Sensor Data vs Predicted



Lancaster Site PM 2.5 Observed Sensor Data vs Predicted



48-hour Prediction Accuracy	Sensor Location
94%	Downtown LA
95%	Long Beach
91%	Lancaster
91%	Glendora
93%	Santa Clarita
93%	Reseda
95%	Long Beach – Rt 710

Predicting **Ozone** Based on Satellite, Ground, Meteorological, and Wildfire/Smoke Data

48-hour prediction Accuracy	Sensor Location
93.53%	Downtown LA
95.90%	Long Beach
91.25%	Santa Clarita
88.19%	Reseda
86.23%	Lancaster
87.35%	Glendora
91.45%	Westchester
87.49%	Pico Rivera
90.04%	Compton
92.87%	Pasadena
93.10%	West LA
92.13%	Azusa
90.59%	Pomona

Accuracy	Frame #
91%	Frame 1: 2 days in future
89%	Frame 2: 4 days in future
86%	Frame 3: 6 days in future
84%	Frame 4: 8 days in future
80%	Frame 5: 10 days in future

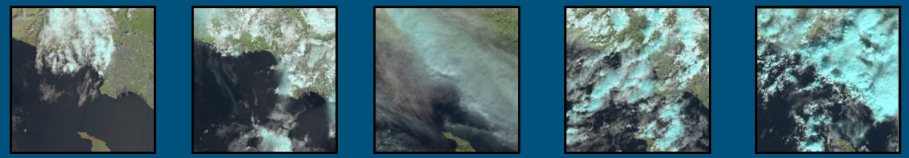
- Satellite observations NASA MODIS: 1 km x 1 km/pixel
- Ground-based sensors (13 in L.A. County), hourly
- Wildfire/Smoke data from NASA MODIS, MERRA-2,
- Meteorological data

Predicting $\underline{NO_2}$ Based on Satellite, Ground, Meteorological, and Wildfire/Smoke Data

Input data

- Satellite observations NASA MODIS:
  - 1 km x 1 km/pixel
- Ground-based sensors (13 in L.A. County), hourly
- Wildfire/Smoke data from NASA MODIS, MERRA-2
- Meteorological data

Accuracy	Frame #
87.62%	Frame 1: 2 days in future
84.15%	Frame 2: 4 days in future
82.38%	Frame 3: 6 days in future
79.06%	Frame 4: 8 days in future
72%	Frame 5: 10 days in future

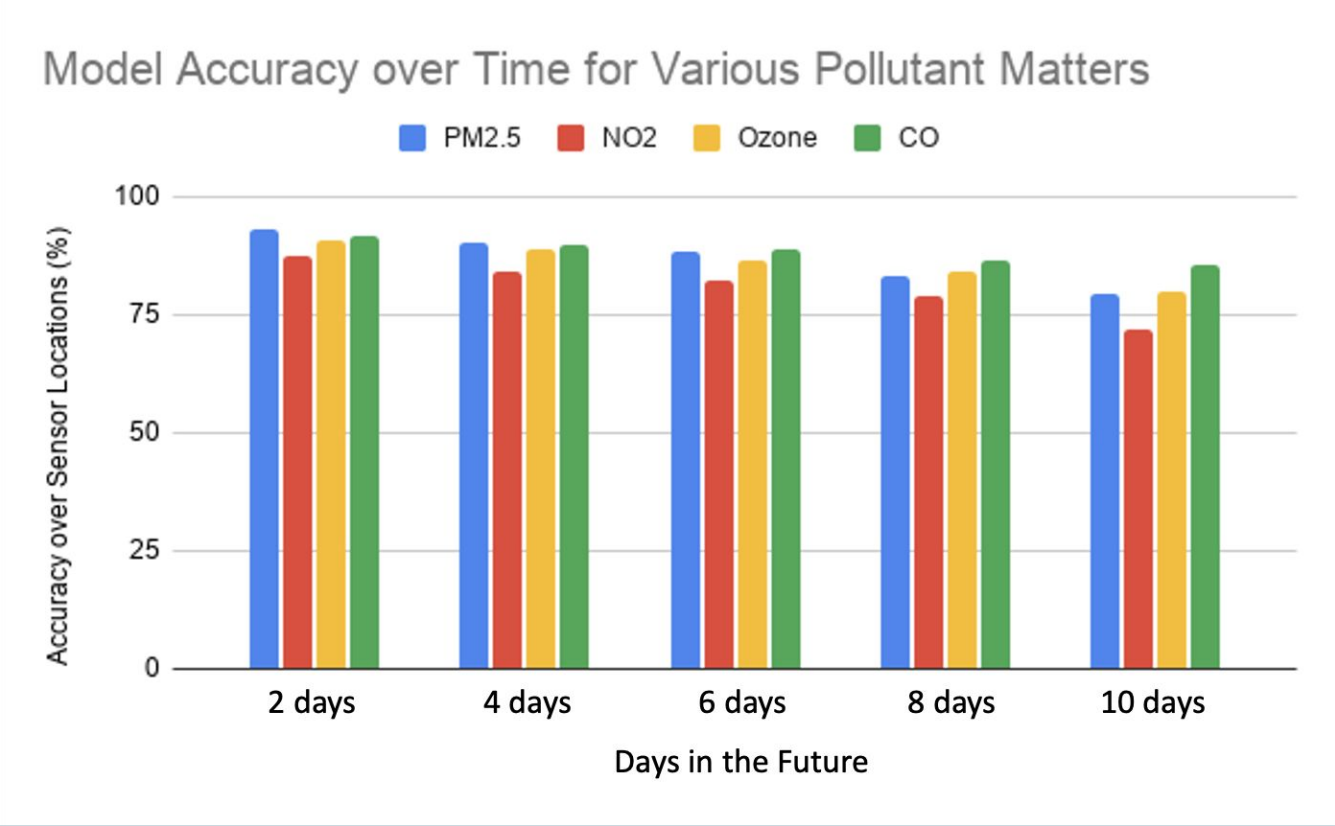


Predicting $\underline{NO_2}$ Based on Satellite, Ground, Meteorological, and Wildfire/Smoke Data

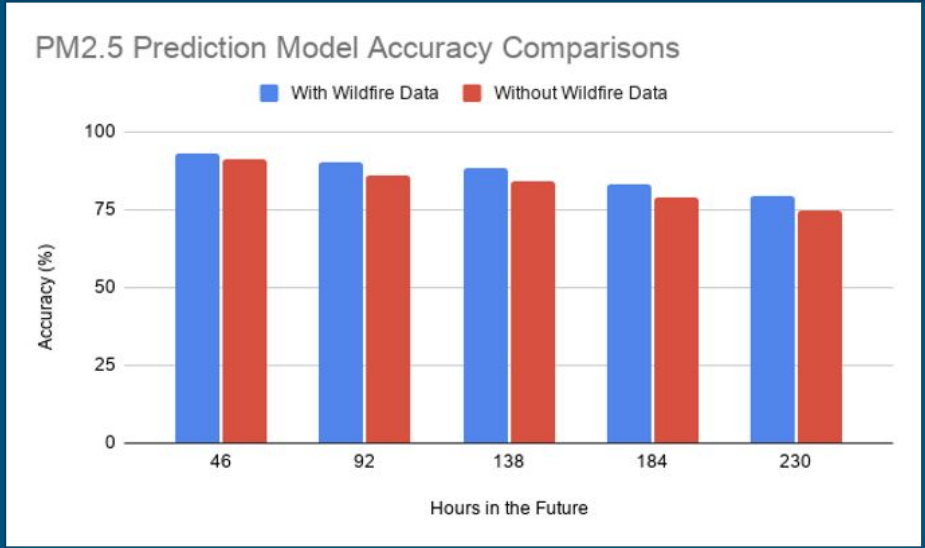
48-hour prediction Accuracy	Sensor Location
93%	Downtown LA
91%	Long Beach
91%	Santa Clarita
89%	Reseda
87%	Lancaster
88%	Glendora
91%	Westchester
91%	Pico Rivera
95%	Compton
92%	Pasadena
90%	West LA
92%	Azusa
92%	Pomona

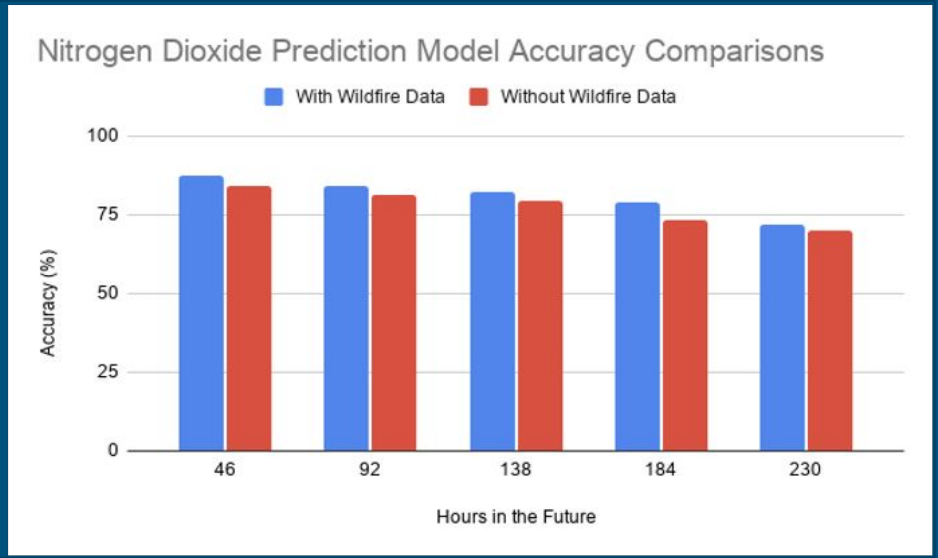
Accuracy	Frame #
87.62%	Frame 1: 2 days in future
84.15%	Frame 2: 4 days in future
82.38%	Frame 3: 6 days in future
79.06%	Frame 4: 8 days in future
72%	Frame 5: 10 days in future

Prediction Results: Model Accuracy for Prediction of Various Pollutant Matters in Future



Model Comparisons: Effect of Wildfire/Smoke Data





Summary

- Team meets regularly and connects to new partners
  - AQMD
  - Propeller Health
  - OpenAQ
  - SmartAirLA
  - SafeCast
  - Southern California Asthma Association
- Fine tuning ML model options
- Close coordination with other AIST partners
  - NASA data standards
  - Analytics Data Framework development
- Already engaging global cities
- Launched citizen science data collection with Mayor on Earth Day
- Participating in ESIP Air Quality Community of Practice
- Held local L.A. Community Air Quality workshop

Partners

- Public
  - City of Los Angeles
  - NASA/JPL
  - Southern California Air Quality Management District
- Private
  - OpenAQ
  - SmartAirLA
  - SafeCast
- Academic
  - California State University, Los Angeles
  - LA Data Science Federation
- Organizations
  - Mayor Garcetti leads the C40 Cities
  - Climate Mayors
  - Pacoima Beautiful